Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world

Travis M. Seale-Carlisle[1], Melissa F. Colloff[1], Heather D. Flowe[1], William Wells[2], John T. Wixted[3] & Laura Mickes[4]

[1]University of Birmingham, [2]Sam Houston State University, [3]University of California, San Diego, [4]University of Bristol

Author Note

Correspondence concerning this article should be addressed to Travis Seale-Carlisle (T.M.Seale-Carlisle@bham.ac.uk) or Laura Mickes (laura.mickes@bristol.ac.uk).

Abstract

The criminal justice system should consider the confidence an eyewitness expresses when making an identification at the time the initial lineup procedure is conducted. High confidence expressed at this time typically indicates high accuracy in the identification. Because the suspect identification – not filler identifications or no identifications – matters most in the court of law, confidence-accuracy characteristic (CAC) analysis provides information most relevant to stakeholders. However, just as high confidence identifications indicate high accuracy, fast identifications may also indicate high accuracy. We tested whether a new technique that is similar to CAC analysis, called response time-accuracy characteristic (RAC) analysis, could inform stakeholders about the likely accuracy of an identification while usefully summarizing response time data. We argue this is the case in the lab and in the real world. Furthermore, CAC and RAC results are not completely redundant so both, considered together, are useful to the criminal justice system.

Keywords: Eyewitness identification, confidence, response time, simultaneous lineup, police lineup, response-time accuracy characteristic analysis

General Audience Summary

During a police investigation, an eyewitness may be presented with a lineup. If the suspect is identified, and if that suspect becomes the defendant in a court of law, judges and jurors should know the likelihood that an eyewitness's identification of the defendant is accurate. One indicator of accuracy is confidence. That is, if an eyewitness expresses high confidence in the identification of the suspect during the initial identification procedure (as opposed to a later time, such as in a courtroom during the trial proceedings), then the identification is more likely to be accurate than if low confidence is expressed. Another indicator of accuracy may be the time it takes for an eyewitness to identify the suspect during the initial identification procedure. That is, if an eyewitness makes an identification of the suspect quickly, is that identification more likely to be accurate than if an identification is made slowly? Whether this is the case and whether confidence and response time could each provide useful independent information for judges and jurors were the main questions that we investigated in this research. For both eyewitness participants in the lab and real eyewitnesses in the field, confidence and response time both provide information about the accuracy of the identifications. Moreover, they provide partially non-overlapping information. Thus, we suggest that, where possible, judges and jurors should be provided with both kinds of information to help them determine the likelihood that defendants are guilty.

Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world

To increase the value of eyewitness identification evidence, the criminal justice system should rely on strong indicators of accuracy. Consider, for example, the information provided by eyewitnesses during the administration of lineup procedures. Lineups are composed of the police suspect (who is innocent or guilty) and several fillers (who are known to be innocent and who resemble the suspect or match the description of the perpetrator). Relying solely on whether or not the suspect was identified, although common, ignores important indicators of accuracy. What indicators should be considered?

Confidence and response times are both known to be strong indicators of accuracy (if and only if gathered during the administration of the initial lineup procedure). Identifications made with high confidence are typically high in accuracy, whereas identifications made with low confidence are less so (e.g., Brewer & Wells, 2006; Carlson et al., 2016; Grabman, Dobolyi, Berelovich, & Dodson, 2019; Juslin, Olsson, & Winman, 1996; Semmler, Dunn, Mickes, & Wixted, 2018; Wilson, Seale-Carlisle, & Mickes, 2018). Likewise, identifications made quickly are higher in accuracy than identifications made slowly (e.g., Brewer, Caon, Todd, & Weber, 2006; Brewer & Wells, 2006; Dobolyi & Dodson, 2018; Dodson & Dobolyi, 2016; Dunning & Perretta, 2002; Flowe & Cottrell, 2011; Sauerland & Sporer, 2009; Smith, Lindsay, & Pryke, 2000; Sporer, 1992; 1993; Weber, Brewer, Wells, Semmler, & Keast, 2004).

How to best convey these empirical outcomes to decision-makers in the criminal justice system is an important consideration. For applied purposes, where suspect identifications are usually the focus, Mickes (2015) recommended conducting confidence accuracy characteristic (CAC) analysis to assess the reliability of the identification. CAC analysis entails computing suspect ID accuracy separately for every level of confidence. For studies that use 6-person lineups with no designated innocent suspect, CAC is given by

$$CAC = \frac{CID_{conf}}{CID_{conf} + FID_{conf}/6}$$

where CID_{conf} is the number of suspect IDs made with a particular level of confidence from target-present lineups, FID_{conf} is the number of filler IDs made with that same level of confidence from target-absent lineups, which is divided by the number of people in the lineup (6 in this example).

CAC analysis is particularly helpful to judges and jurors who have to determine the culpability of identified defendants because it is straightforward (i.e., a simple proportion correct) and is directly relevant to the issue at hand (i.e., the accuracy of the suspect ID that occurred in this investigation, without regard for filler IDs and lineup rejections). The CAC measure, which is also known as *positive predictive value* (PPV) for the equal base-rate scenario,[1] is a convenient way to understandably summarize the information value of a suspect ID made with a certain level of confidence.

Wixted and Wells (2017) conducted CAC analysis on data from 20 experiments, and the results unambiguously showed that confidence is highly informative of accuracy. In 15 experiments, the same confidence rating scale was used which afforded the opportunity to average the data across experiments. CAC analysis revealed that identifications made with low confidence were less than 70% correct and identifications made with high confidence were ~97% correct. It is clear that initial confidence adds to the evidentiary value above and beyond simply knowing only whether the defendant was identified from a lineup. Thus,

[1] Equal base rate means that half of the lineups contain the guilty suspect and half of the lineups contain the innocent suspect.

arming judges and jurors with this information will help them to assess the likelihood of the defendant's guilt, which depends on the level of confidence the eyewitness expressed in the suspect identification that occurred early in the police investigation (i.e., during the initial lineup procedure).

For similar reasons, an argument can be made that response time data should also be presented to judges and jurors and in a manner analogous to a CAC plot. In fact, in some investigations, in addition to whether the suspect, a filler, or no identification was made, response time is the only information about the likely accuracy of the initial ID that the judges and jurors have. To communicate the information values of the response time associated with a suspect ID, the equation we propose here is similar to the equation used in CAC analysis. However, confidence is replaced by response time. We refer to this analytical approach as response time-accuracy characteristic (RAC) analysis, which for a 6-person lineup is given by

$$RAC = \frac{CID_{RT}}{CID_{RT} + FID_{RT}/6}$$

where *RT* is response time, CID_{RT} is the number of suspect IDs made within a particular response time from target-present lineups (e.g., number of suspect IDs made with an $RT < 5$ seconds), FID_{RT} is the number of filler IDs made within a particular response time from target-absent lineups, which is divided by the number of people in the lineup (6). Like a CAC plot, an RAC plot provides a measure of PPV for the equal base-rate scenario. Thus, RAC provides another possible measure of reliability because it is an estimate of the likelihood that the suspect identified from the lineup is guilty.

To test the hypotheses that both CAC and RAC can provide information about accuracy, we analysed data from an online lab-based forensically-relevant study and, separately, analysed data from a police department field study (Wixted, Mickes, Dunn, Clark, & Wells, 2016). In both studies, we tested participant eyewitnesses' or real eyewitnesses' memory of the perpetrator on fair 6-person simultaneous lineups (i.e., the suspect did not stand out among the fillers). Simultaneous lineups were chosen because they yield better discriminability (i.e., the ability to distinguish innocent from guilty suspects) than sequential lineups in the lab (e.g., Dodson & Dobolyi, 2013; Gronlund et al., 2012; Mickes, Flowe, & Wixted, 2012; Seale-Carlisle, Wetmore, Flowe, & Mickes, in press) and in the field (Amendola & Wixted, 2015; Wixted et al., 2016). This is why we (e.g., Seale-Carlisle et al., in press) and others (e.g., Meisters, Diedenhofen, & Musch, 2018) have recommended that the police use simultaneous lineups. For this applied reason, our investigation focuses on performance on simultaneous lineups (i.e., sequential, showups, and live lineups are outside of the scope of this paper).

Based on prior research, we predict that identifications made with high confidence will be more accurate than those made with lower confidence, and identifications made quickly will be more accurate than slower identifications. The specific questions of interest were 1) is this true of not only lab data but also of data collected in the field from real eyewitnesses? 2) Are the results similar for both lab and field studies? 3) Are the response time (latency) data usefully summarized in terms of RAC?

Study 1: Laboratory Study

Study 1 was a forensically-relevant study in which each participant-eyewitness was presented with a video of a mock crime that contained one target and was then tested using either a target-present or a target-absent 6-person simultaneous lineup. To assess the predictive value that confidence has in suspect identification accuracy, participants made a confidence judgment after making an identification. To assess the predictive value that response times have in identification accuracy, response time from the onset of the lineup to the identification was recorded.

Method

Participants

We aimed to recruit 1,000 participants from Amazon Mechanical Turk (www.mturk.com) and went slightly over target (*N* = 1,046). The average age of the participants was 34.33 years (*sd* = 11.66). Participants were 46% women and 53% men (1% preferred not to state) and 26% Asian, 6% Black, 5% Hispanic, 2% Native American, 59% White, and 2% Other (1% preferred not to state). Participants were randomly assigned to a target-present (*n* = 531) or target-absent (*n* = 515) lineup. Ethical approval was granted by the University of California, San Diego Institutional Review Board project number 121186.

Materials

Video. A young adult White male acted as the target in a 28 s video of a mock crime of vandalism. He was graffitiing a wall outdoors when he noticed a witness and walked off past the witness. The video was filmed from the point of view of the witness. The front of the target's face was clearly shown for 8 s.

Lineups. Target-present lineups contained the target and 5 fillers and target-absent lineups contained 6 fillers. The target and fillers were randomly positioned in the lineup for each participant. Fillers matched the description of the target based on descriptions from 15 participants (not the participants who took part in the experiment proper). These participants viewed the video and answered questions about the target's ethnicity, sex, age, physical build, and hair color. The average or modal responses, as required, were entered into the

Florida Department of Corrections database and based on that search, 50 images of individuals were selected. Fillers were randomly selected from this pool to be displayed in lineups for each participant. There was no designated innocent suspect in target-absent lineups.

Procedure

Participation took place online, but not on mobile devices. Participants were instructed via screen presentation and voiceover audio that they would watch a brief video and that they should pay close attention because they would be asked questions about it. After watching the video, and after a 5-minute distractor game of Tetris, participants were tested on their memory for the target in the lineup.

Before the lineup was presented, participants were instructed via screen presentation and voiceover audio that they would see lineup members at the same time, that the person from the video may or may not be in the lineup, and to select him if he is in the lineup and select the “not present” option if he is not in the lineup. Participants made their decision by clicking on a lineup member’s image or the “not present” button. All of the images were preloaded before presentation and were displayed at the same time. Response time was measured from the time that the members were displayed until the participant made their selection. After their selection was made, the confidence scale appeared, and they were prompted to rate their confidence on an 11-point scale (0 = just guessing and 100 = absolutely certain, binned in tens) using a drop-down menu. Participants next provided demographic information (i.e., age, sex, and ethnicity), answered a multiple-choice validation question (“What crime was committed in the video?”), and were debriefed.

Results and Discussion

Participants who incorrectly answered the validation question were excluded from analyses (*n* = 156). Of the 890 participants remaining, 443 were tested on a target-absent lineup and 447 were tested on a target-present lineup. Because there was no designated innocent suspect, the false suspect IDs were estimated by dividing by the number of fillers in the target-absent lineups (6). The overall false ID rate was 0.05 and the correct ID rate was 0.62. Discriminability measured by partial area under the receiver operating characteristic curve (pAUC) was 0.12 (using a false ID cutoff of 0.72) and by *d′* was 1.98.[2]

Figure 1A shows the CAC results. In a CAC plot, suspect ID accuracy is plotted on the y-axis and confidence on the x-axis. Suspect IDs (from target-present and target-absent lineups) were binned into low (ratings of 0-60), medium (ratings of 70-80) and high (ratings 90-100) levels of confidence (e.g., Mickes, 2015). Identifications made with high confidence were higher in accuracy than identifications made with medium confidence, which were higher in accuracy than identifications made with low confidence. The sizes of the symbols in the figures represent the number of identifications at a given level of confidence (or response time) bin relative to the number of identifications given at other levels of confidence (or response time) bins (Seale-Carlisle et al., in press). As shown by the relative size of the points, the majority of identifications were made with medium confidence, fewer with low confidence, and even fewer with high confidence. The CAC plot tells a clear story that is useful for judges and jurors: confidence is indicative of accuracy (e.g., Mickes, 2015; Wixted & Wells, 2017).

[2] pAUC was conducted using the target-absent filler ID rate not the estimated false ID rate (e.g., Wilson et al., 2018), and *d′* was estimated using an approximation described in Mickes, Moreland, Clark and Wixted (2014).

Figure 1B shows the RAC results. In a RAC plot, suspect ID accuracy is plotted on the y-axis and response time on the x-axis. To create this figure, responses were binned into four response times (< 5 seconds, 5-15 seconds, 16-30 seconds, and > 30 seconds).[3] The data were binned in this manner, because it is a feasible way for police to collect response time in the real world (the field data discussed below were originally collected in 5-second bins). As response time increased, accuracy decreased. As shown by the relative size of the points, the majority of responses were made between 5 and 15 seconds, quite a few were made in less than 5 seconds, and very few responses took longer than 15 seconds. As with confidence, response time is indicative of accuracy. Moreover, as with a CAC plot for confidence, an RAC plot would appear to be a useful way to convey the information-value of response time to decision-makers in the legal system.

It is possible that confidence and response time provide completely redundant indicators of accuracy. To investigate this issue, we compared high-confidence suspect IDs made in less than 5 s to those made between 5 s and 15 s to those made after 15 s. The data were combined in this manner to have enough observations for meaningful comparisons. We also combined medium- and low-confidence suspect IDs for the same reason. Table 1 shows that, in fact, confidence and response time provide partially independent information. For high-confidence suspect IDs, PPV, while high, was nevertheless lower for long-latency decisions (> 15 s) than for short-latency decisions (< 5 s) (Table 1A). The same trend was apparent for the combined medium- and low-confidence IDs (Table 1B). To assess the significance of these trends, we used the Fisher's Exact test (because some cells had fewer

[3] Response times were rounded to the nearest second.

than 5 observations).[4] No differences were significant at $p < .05$. However, because there were so few estimated false IDs, especially made with high confidence, we conducted the analysis with target-absent filler IDs instead of estimated false IDs. Note that the target-absent lineup consists of 6 fillers, which is analogous to a real police lineup that (unbeknownst to the police) contains an innocent suspect. From the point of view of the witness, the lineup simply consists of 6 people who match the description of the perpetrator but who do not match memory (i.e., they are all description-matched fillers from the witness's perspective). In a lab study, we can therefore statistically analyse all of the target-absent filler IDs because, if the lineup is fair, what is statistically true of filler IDs must also be statistically true of innocent suspect IDs. When the data were analyzed in this way the association between the type of ID (ID: target-present suspect ID vs. target-absent filler ID) and response time (< 5 s vs. 5-15 s vs. > 15 s) was significant for high confidence IDs, $\chi^2(2, 156) = 9.18$, $p = .010$, and medium- and low-confidence IDs, $\chi^2(2, 247) = 10.75$, $p = .005$. Thus, within a level of confidence, slower IDs were less accurate than faster IDs. The data are available here: https://figshare.com/s/07f5f6af36d00a8380b1.

Study 2: Police Department Field Study

In Study 1 (a lab-based study), both confidence and response time were indicative of suspect identification accuracy. Whether this is also true in the real world is of applied interest. In a police department field study in Houston, Texas, real eyewitnesses to crimes were presented with 6-person simultaneous or sequential lineups, and if identifications were made, they provided expressions of confidence (Wixted et al., 2016). Unlike in the lab, in the

[4] To conduct the Fisher's Exact test, the estimated false IDs were rounded up to the nearest whole number.

real world, whether or not the suspect is actually the perpetrator is unknown. Nonetheless, confidence was predictive of accuracy because the majority of filler identifications was made with low confidence, and the majority of suspect identifications was made with high confidence (Wixted et al., 2016). Because the lineups (a) involved only strangers, (b) were fair, and (c) were administered in double-blind fashion (i.e., the person administering the lineup did not know who was the suspect), the only way to identify the suspect at greater-than-chance accuracy was to rely on memory. That fact makes it possible to use a model of recognition memory to estimate suspect ID accuracy. Using either a simple high-threshold model or a more complex signal detection model, Wixted et al. found that confidence was strongly related to accuracy (in a manner similar to that of lab data). Because the two approaches to estimating suspect ID accuracy yielded nearly identical values, we use the simpler (purely algebraic) threshold model to estimate suspect ID accuracy from the response time data here.[5]

Results and Discussion

According to the threshold model, the probability of a correct suspect ID from a target-present lineup is equal to the probability that a witness recognizes the perpetrator (p) plus the probability that a witness who does not recognize the perpetrator makes a random guess that happens to land on the guilty suspect. The latter probability is equal to the probability that the witness makes a guess (g) times 1/6, or $g(1/6)$. The probability of an incorrect suspect ID from a target-absent lineup is equal to the probability that the witness makes a random guess that happens to land on the guilty suspect, which is equal to $g(1/6)$.

[5] The response time data were not previously published.

Similar equations correspond to the probability of making filler IDs from target-present and target-absent lineups. Using the observed number of suspect and filler IDs from both target-present and target-absent lineups, one can algebraically obtain estimates of *p* and *g* for each level of confidence (or for each response time bin) and then use those estimates to estimate suspect ID accuracy (Wixted et al., 2016). Note that face recognition memory is almost certainly not an all-or-none process (as the threshold model assumes), but the virtue of the threshold model is the simplicity of the equations that can be used to estimate suspect ID accuracy.

As shown in Figure 2A, when memory is tested using simultaneous lineups (*N* = 187), the threshold model estimates indicate that confidence is predictive of accuracy.[6] That is, identifications made with high confidence had higher PPV than identifications made with medium confidence, which, in turn, had higher PPV than identifications made with low confidence. This pattern is consistent with the lab-based CAC results presented in Study 1 and with many other lab-based results (e.g., Wixted & Wells, 2017).

The relative size of the points indicates that most suspect identifications were made with high confidence, with many fewer identifications being made with medium or low confidence. Though not shown here, as noted by Wixted et al. (2016), the opposite pattern was obtained for filler IDs, with most filler IDs being made with low confidence and many fewer filler IDs made with high confidence. This pattern was observed despite the fact that, in a 6-person lineup, there are many more opportunities for filler IDs to occur than suspect IDs. For simultaneous lineups, high-confidence suspect IDs far outnumbered high-confidence filler IDs (49 to 9), whereas low-confidence suspect IDs were far outnumbered by low-confidence

[6] A 50% base rate was assumed. For CAC estimates of other base rates see Wixted et al. 2016.

filler IDs (6 to 24). On their own (i.e., without the aid of a model of recognition memory), those opposing patterns indicate that confidence is a strong predictor of accuracy even in real world data.

In addition to recording eyewitnesses' expressions of confidence, the administrators also recorded response time in 5 s bins for the majority of these simultaneous lineups (*n* = 141). We conducted RAC analysis (using the threshold model to again estimate suspect ID accuracy), and the results are shown in Figure 2B. The longer it took to make an identification, the lower the suspect identification accuracy. On the other hand, identifications made in less than 5 s were highly accurate. This pattern is consistent with the RAC results in Study 1. In addition, the relative size of the points shows that most of the suspect identifications were made within 5 s and very few identifications were made after 30 s.

To investigate whether confidence and response time provide completely redundant indicators of accuracy in the police department field study data, we compared high-confidence IDs made in less than 5 s to those made after 15 s and medium- and low-confidence IDs made in less than 5 s to those made after 15 s. Combining the data in this manner allowed for enough observations for a meaningful test. Table 2 shows that confidence and response time provide partially independent information (similar to the lab data in Table 1). P(SID|ID) is the probability that the ID landed on the suspect given that an ID was made, and PPV is the estimated suspect ID accuracy according to the threshold model. For high-confidence IDs, PPV, while always high, was nevertheless lower for long-latency decisions than for short-latency decisions (Table 2A). According to the Fisher's Exact test, the association between the type of ID (suspect vs. filler) and response time (< 5 s vs. > 15 s) was significant for high confidence IDs, $p = 0.019$. The same trend was apparent for the combined medium- and low-confidence IDs (Table 2B), but the association was not significant ($p > 0.05$). Thus, there are indications that within a level of confidence (significantly so for high confidence IDs), slower IDs are less accurate than faster IDs.

General Discussion

In two studies – a lab-based study involving witnesses to a mock crime and a police department field-based study conducted with the Houston Police Department involving witnesses to real crimes – individuals were administered fair 6-person simultaneous lineups to test the hypotheses that confidence and response time both provide useful indicators of accuracy. Both hypotheses were supported. In agreement with much prior work, identifications made with high confidence were highly accurate and were more accurate than medium- and low-confidence identifications. In addition, identifications made quickly (i.e., under 5 seconds) were highly accurate and were more accurate than identifications that were made more slowly. Thus, also in agreement with prior work, these results show that response time, much like confidence, is an indicator of accuracy. Importantly, much like with confidence, we have shown that response time can be simply displayed to legal decision-makers using RAC analysis.

The results reported here are important, but they are not surprising. The confidence results are consistent with previous applied (e.g., Brewer & Wells, 2006; Wixted & Wells, 2017) and basic research (e.g., Mickes, Hwe, Wais, & Wixted, 2011; Mickes, Wais, & Wixted, 2009; Roediger & DeSoto, 2014; Tekin & Roediger, 2017; Weber & Brewer, 2003; Wixted & Mickes, 2010). The response time results are also consistent with previous applied (e.g., Brewer & Wells, 2006; Dobolyi & Dodson, 2018; Dunning & Perretta, 2002; Grabman et al., 2019; Sauerland & Sporer, 2009; Sporer, 1992; 1993; 1994) and basic research (e.g., Norman & Wickelgren, 1969; Ratcliff, 1978; Ratcliff & Murdock, 1976). Whereas confidence has previously been shown to provide an important indicator of accuracy for real eyewitnesses, this is the first time that response time data has also been found to provide a reliable indicator of accuracy for real eyewitnesses tested using police lineups. Moreover, and critically, we propose here for a new method for effectively communicating the information value of response times associated with suspect IDs to decision-makers, namely, RAC analysis[7].

This approach differs from – and we would argue is much better than – previous approaches that have focused on identifying an optimal response time boundary (i.e., response time cut-off that best separates accurate from inaccurate responses). Indeed, the relationship between accuracy and response time appears to be continuous (Weber et al., 2004); our data do not provide any support for the 10-12-second cut-off rule or any similar categorical rule (Dunning & Perretta, 2002; Sauerland & Sporer, 2009; Weber et al., 2004). Thus, reporting the information-value of response time in a way that reflects its continuous nature, as RAC analysis does, seems like an ideal approach.

Others have assessed confidence, response time, and accuracy by fitting a linear mixed effects model (e.g., Dodson & Dobolyi, 2018; Grabman et al., 2019). CAC and RAC analyses do not require any model fitting, which requires various assumptions to be met. CAC and RAC analyses also directly respond to the needs of the criminal justice system by providing information about the suspect's guilt. Figure 3 shows the CAC (Figure 3A) and RAC (Figure 3B) plots constructed from the data reported in Grabman et al. In this experiment,

[7] For theory development, filler IDs should be included in the CAC analyses (Mickes, 2015) and RAC analysis.

participants were shown a list of faces and then tested on lineups either 5 minutes or one day later. Both analyses show that identifications that were made with higher confidence and faster were associated with higher accuracy than those made with lower confidence and slower. These data reinforce the message that confidence and response times are indicative of accuracy and CAC and RAC analyses readily show it.

To increase the value of eyewitness evidence, researchers have long recommended that the police collect witnesses' expressions of confidence in their identifications during the administration of the initial lineup procedures (e.g., Brewer & Palmer, 2010; Juslin et al., 1996). Wixted, Mickes, Clark, Gronlund, and Roediger (2015), made what is perhaps the most emphatic claim:

> …because eyewitness confidence on an initial test of memory is diagnostic of guilt (whereas eyewitness confidence on a later memory test may not be), it is important that the police record the initial level of confidence expressed by an eyewitness. (p. 523).

The results reported here lead us to make a similar recommendation about response time: Because the time that it takes for an eyewitness to make an identification is diagnostic of guilt of an initial identification, it is important that the police record the time between presentation of the lineup and the identification.

Critically, confidence and response time, considered together, provided more information than either one considered alone. For a given level of confidence, accuracy was higher when the ID was made quickly compared to when it was made slowly. During investigations, the police should be made aware of how confident an eyewitness is in the suspect identification and the time the suspect identification took. This will inform police investigators in their assessments of eyewitness selections. Furthermore, judges and jurors should be made privy to that information to inform them in their assessments of defendant culpability.

With regard to the theoretical implications of our findings, it seems that a standard signal detection model may not fully account for the data shown in Table 1. A rule-of-thumb derived from signal detection models is that high confidence IDs will be made quickly, whereas low-confidence IDs will be made more slowly. More specifically, a high-confidence ID occurs when the memory signal associated with a lineup member is strong (i.e., it falls far above the decision criterion for making an ID). The more the memory signal exceeds the criterion, the higher the confidence and the faster the response time. However, our findings suggest that even within a given level of confidence, fast IDs are more reliable than slow IDs. Because signal detection models do not formally account for response time, and sequential sampling models do (but generally do not account for confidence), a hybrid signal-detection/sequential-sampling model may ultimately be best suited to account for these results.

A sequential sampling model applied to lineups assumes that information begins to accumulate the moment the photos are presented to the witness. The noisy information accumulation process continues until the cumulative signal reaches an upper threshold (at which point an ID would be made) or a lower threshold (at which point the lineup would be rejected). The time required to reach a threshold represents the RT. A sequential sampling model proposed by Pleskac and Busemeyer (2010) further assumes that evidence is then accumulated for an additional fixed period of time, and that evidence variable is modelled as a standard signal detection process, yielding a confidence rating. Alternatively, a sequential-sampling model like RTCON2 might be considered (Ratcliff & Starns, 2009; 2013). RTCON2 is a variant of the influential diffusion model (Ratcliff, 1978), one that incorporates a signal-detection-like process to model both confidence and RT at the moment a threshold is reached. Conceivably, RTs from lineup memory tasks would offer a new way to differentially evaluate the predictions made by models such as these.

For applied purposes, the information provided by confidence and response time increases the probative value of eyewitness evidence. We therefore propose that when police investigators are assessing the evidence of a crime, the results of CAC analyses and RAC analyses should be considered. We also propose that when judges and jurors are assessing the culpability of defendants based on eyewitness evidence, the results of CAC analyses or RAC analyses (or both, depending on the nature of the eyewitness evidence) should be provided and considered. With such information in hand, these decision-makers would be better equipped to make decisions about the likelihood of guilt because the CAC and RAC results speak directly to the question of interest (what is the probability that the identified suspect is guilty?) and are presented in a way that is easy to understand (e.g., identifications that are made with high confidence and identifications that are made quickly are typically over 95% accurate).

Author Contributions

All authors conceived of the study idea and developed the concept. The first, fourth, and sixth authors designed the studies. The first author programmed Study 1 and collected the data, and the fourth author was involved in data collection for Study 2. The first, fifth, and sixth authors were involved in data analysis and interpretation. The first and sixth authors wrote the initial draft of the manuscript and all authors were involved in editing and revising the manuscript.

References

Amendola, K. L. & Wixted, J. T. (2015). Comparing the diagnostic accuracy of suspect identifications made by actual eyewitnesses from simultaneous and sequential lineups in a randomized field trial. *Journal of Experimental Criminology, 11,* 263-*284.*

Brewer, N., Caon, A., Tood, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior, 30,* 31-50.

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology, 15,* 77-96.

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12,* 11-30.

Carlson, C. A., Young, D. F., Weatherford, D. R., Carlson, M. A., Bednarz, J. E., & Jones, A. R. (2016). The influence of perpetrator exposure time and weapon presence/timing on eyewitness confidence and accuracy. *Applied Cognitive Psychology, 30*, 898-910.

Dobolyi, D. G., & Dodson, C. S. (2018). Actual vs. perceived eyewitness accuracy and confidence and the featural justification effect. *Journal of Experimental Psychology: Applied*, *24*, 543-563.

Dobolyi, D. G. & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied, 19*, 345–357.

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and Eyewitness Identifications: The Cross-Race Effect, Decision Time and Accuracy. *Applied Cognitive Psychology*, *30*, 113-125.

Dunning, D. & Perretta, S. (2002). Automaticity and eyewitness accuracy: a 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology, 87,* 951-962.

Flowe, H. D., & Cottrell, G. (2011). An examination of simultaneous lineup decision processes using eye tracking. *Applied Cognitive Psychology, 25*, 443-451.

Grabman, J. H., Dobolyi, D. G., Berelovich, N. L., & Dodson, C. S. (2019). Predicting High Confidence Errors in Eyewitness Memory: The Role of Face Recognition Ability, Decision-Time, and Justifications. *Journal of Applied Research in Memory and Cognition, 8*, 233-243.

Gronlund, S. D., Carlson, C.A., Neuschatz, J.S, Goodsell, C.A., Wetmore, S. A., Wooten, A., & Graham, M. (2012). Showups versus lineups: An evaluation using ROC analysis. *Journal of Applied Research in Memory and Cognition*, *1*, 221-228.

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1304-1316.

Mickes, L. (2015). Receiver Operating Characteristic Analysis and Confidence-Accuracy Characteristic Analysis in Investigations of System Variables and Estimator Variables that Affect Eyewitness Memory. *Journal of Applied Research in Memory and Cognition, 4,* 93-102.

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver Operating Characteristic Analysis of Eyewitness Memory: Comparing the Diagnostic Accuracy of Simultaneous vs. Sequential Lineups. *Journal of Experimental Psychology: Applied, 18,* 361-376.

Mickes, L., Hwe, V., Wais, P. E., & Wixted, J. T. (2011). Strong memories are hard to scale. *Journal of Experimental Psychology: General, 140,* 239-257.

Mickes, L., Moreland, M. B., Clark, S. E. & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute *d'*, not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition, 3*, 58-62.

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual process theories of recognition memory. *Psychological Science, 20,* 509-515.

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology, 6,* 192–208.

Pleskac, T. J. & Busemeyer, J. R. (2010). Two Stage Dynamic Signal Detection: A Theory of Choice, Decision Time, and Confidence. *Psychological Review, 117*, 864-901.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59-108.

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83, 190–214.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological review*, *116*(1), 59.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological review*, *120*(3), 697.

Roediger, H. L., & DeSoto, K. A. (2014). Confidence and memory: Assessing positive and negative correlations. Memory, 22(1), 76-91.

Sauerland, M., & Sporer, S. L. (2009). Fast and confidence: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied, 15,* 46-62.

Seale-Carlisle, T. M., Wetmore, S., Flowe, H., & Mickes, L. (in press). Designing police lineups to maximize memory performance. *Journal of Experimental Psychology: Applied.*

Semmler, C., Dunn, J., Mickes, L., & Wixted, J. T. (2018). The Role of Estimator Variables in Eyewitness Identification. *Journal of Experimental Psychology: Applied, 24*, 400-415.

Smith, S. M., Lindsay, R. C. L., & Pryke, S. (2000). Postdictors of eyewitness errors: Can false identifications be diagnosed in the cross-race situation*? Journal of Applied Psychology, 85*, 542-550.

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology*, *22*, 157–180.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, *78*, 22–33.

Sporer, S. L. (1994). Decision–times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F.Ross, J. D.Read, & M. P.Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300–327). New York: Cambridge University Press.

Tekin, E., & Roediger, H. L. (2017). The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cognitive Research: Principles and Implications, 2,* 49.

Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology, 88,* 490-499.

Weber, N. Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal of Experimental Psychology: Applied, 10,* 139-147.

Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L**.** (2018). The effects of verbal descriptions on performance in lineups and showups. *Journal of Experimental Psychology: General, 147,* 113-124*.*

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological review*, *117*, 1025-1054.

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E. & Wells, W. (2016). The reliability of eyewitness identifications from police lineups. *Proceedings of the National Academy of Sciences, 113*, 304-309.

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70,* 515-526.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18*, 10–65.

Table 1.

Target-present suspect IDs and target-absent estimated suspect IDs made in less than 5 s, between 5 and 15 s, and greater than 15 s for Study 1.

Latency	Target-present suspect ID	Target-absent estimated suspect ID	PPV
High confidence			
< 5 s	56	11/6 = 1.83	0.97
5 - 15 s	66	11/6 = 1.83	0.97
> 15 s	6	3/6 = 0.50	0.92
Medium- and low-confidence			
< 5 s	46	15/6 = 2.50	0.95
5 - 15 s	81	56/6 = 9.33	0.90
> 15 s	22	27/6 = 4.50	0.83

Note. Positive predictive value = PPV is suspect ID accuracy. The target-absent estimated suspect ID column shows the number of filler IDs divided by lineup size of 6.

Table 2.

IDs made in less than 5 s and greater than 15 s for Study 2.

High confidence IDs				
Latency	Suspect ID	Filler ID	p(SID|ID)	PPV
< 5 s	26	2	0.93	0.99
> 15 s	10	6	0.63	0.94
Medium- and low-confidence				
Latency	Suspect ID	Filler ID	p(SID|ID)	PPV
< 5 s	3	3	0.50	0.90
> 15 s	8	28	0.22	0.64

Note. P(SID|ID) is the probability of a suspect ID (SID) given that an ID was made, and positive predictive value (PPV) is the estimated suspect ID accuracy according to the threshold model.

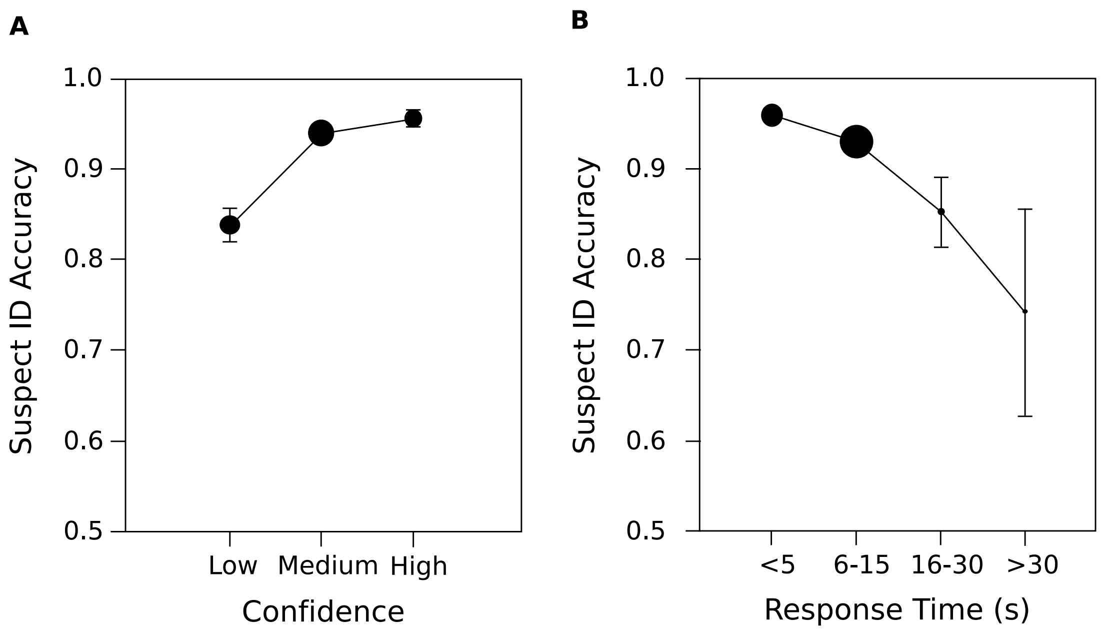

Figure 1. (A) CAC results and (B) RAC results of Study 1. The bars represent standard errors. The size of the symbols represents relative frequencies.

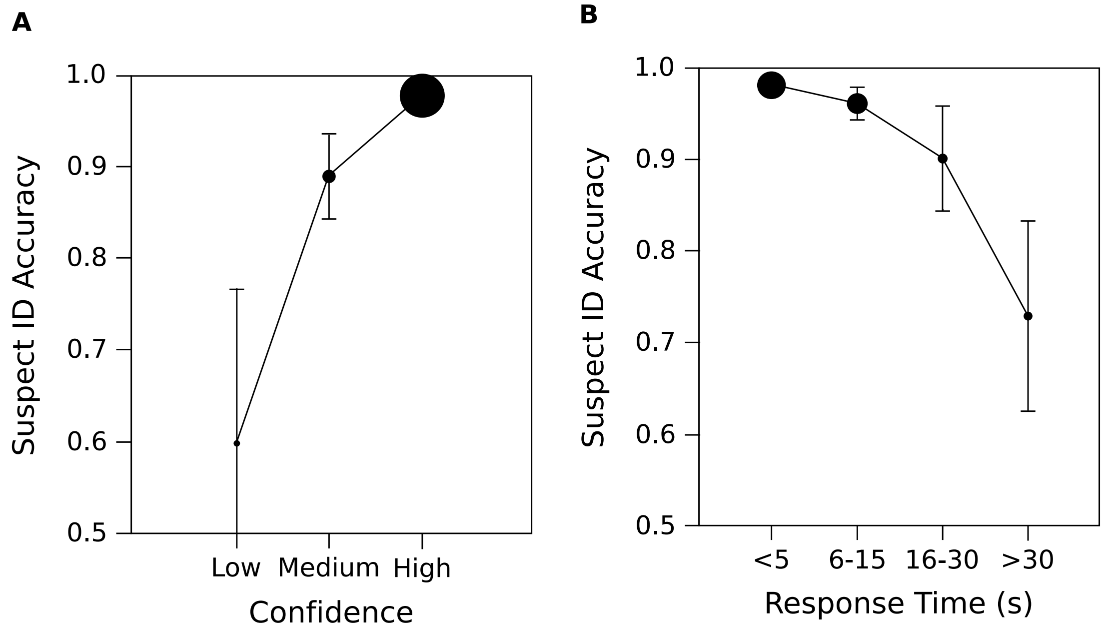

Figure 2. (A) CAC results (assuming a 50% base rate) and (B) RAC data of Study 2. The bars represent standard errors. The size of the symbols represents relative frequencies.

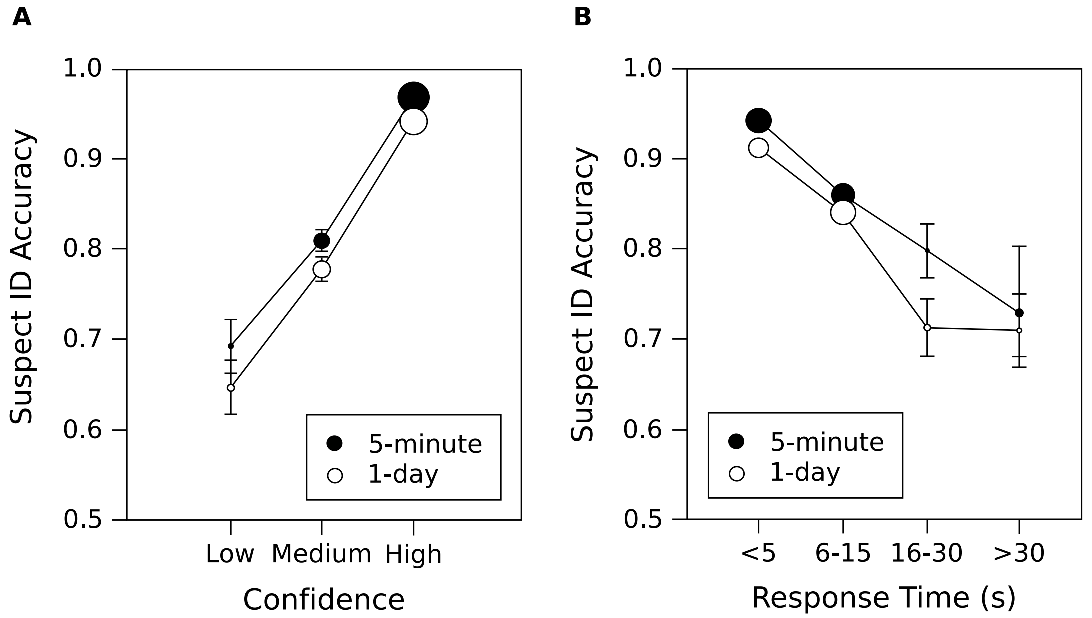

Figure 3. (A) CAC results and (B) RAC data from Grabman et al. (2019). The bars represent standard errors. The size of the symbols represents relative frequencies.